Semi-Nonparametric Modeling and Estimation*

Herman J. Bierens
Pennsylvania State University
Department of Economics and CAPCP†
University Park, PA 16802, USA

April 25, 2009

Abstract

In this paper it will show how unknown density and distribution functions can be modeled semi-nonparametrically via orthonormal series expansions, and how to estimate semi-nonparametric (SNP) models via a sieve estimation approach. As an application I will focus on the mixed proportional hazard (MPH) model with fixed right censoring and unspecified mixing distribution and baseline hazard. I will show how the MPH model with fixed right censoring can be estimated consistently by an integrated method of moments sieve estimation approach. Another application is the first-price auction model, which will also be discussed, but more briefly than the MPH model.

1 Introduction

Semi-nonparametric (SNP) models are models for which the functional form is only partly parametrized and where the non-specified part is an unknown

*This paper has been prepared for a keynote address given at the *Primer Congreso Científico Internacional en Economía y Finanzas,* May 18-22, 2009, in Quito, Ecuador.

†Support for research within the Center for the Study of Auctions, Procurements, and Competition Policy (CAPCP) at Penn State has been provided by a gift from the Human Capital Foundation

function. See for example Chen (2007) and the references therein. In this paper I will only consider the case where this unknown function is modeled via an orthonormal series expansion.

Gallant (1981) was the first econometrician to proposed Fourier series expansions as a way to model unknown functions. See also Eastwood and Gallant (1991) and the references therein. However, the use of Fourier series expansions to model unknown functions has been proposed earlier in the statistics literature. See for example Kronmal and Tarter (1968).

Gallant and Nychka (1987) consider SNP estimation of Heckman's (1979) sample selection model, where the bivariate error distribution of the latent variable equations is modeled semi-nonparametrically using a bivariate Hermite expansion of the error density.

Another example of a SNP model is the mixed proportional hazard (MPH) model proposed by Lancaster (1979), which is an proportional hazard model with unobserved heterogeneity. Heckman and Singer (1984) propose to estimate the distribution function of the unobserved heterogeneity variable by a discrete distribution. Bierens (2008a) and Bierens and Carvalho (2007) use orthonormal Legendre polynomials to model semi-nonparametrically the unobserved heterogeneity distribution of interval-censored mixed proportional hazard models and bivariate mixed proportional hazard models, respectively.

In this paper I will first discuss the MPH model, in section 2, including its nonparametric identification, and the fixed right censoring case. In the latter case the baseline hazard can be left unspecified, together with the mixing distribution, which gives rise to a SNP model with Euclidean parameters for the systematic hazard and two non-Euclidean parameters in the form of unknown distribution functions on the unit interval. Moreover, an integrated conditional moments method will be proposed for the estimation of these parameters. The latter is a new result.

In section 3 the sieve estimation method is reviewed. In section 4 it will be shown how to model densities and distribution functions on the unit interval semi-nonparametrically, and how to construct compact metric spaces of density and distribution functions. The latter are needed as parameter spaces for the two distribution functions involved.

In section 5 it will be shown how the SNP sieve estimation approach can be applied to estimate the value function of a first-price auction model, using a simulated integrated moments method. Section 6 is a technical appendix containing some proofs.

Throughout this paper I will use the following notations and abbreviations. The indicator function will be denoted by $I(.)$: $I(true) = 1$, $I(false) = 0$. The abbreviation "a.s." stands for "almost surely", which means that the result involved holds with probability one. Similarly, the abbreviation "a.e." stands for "almost everywhere", which means that the result involved holds except perhaps on a set with Lebesgue measure zero. Finally, "$\sim$" stands for "is distributed as".

2 The Mixed Proportional Hazard Model

2.1 The general MPH model

The mixed proportional hazard (MPH) model proposed by Lancaster (1979) assumes that the hazard function of a duration T is the product of three factors: a baseline hazard $\lambda_0(t) \geq 0$, which depends on time only, a systematic hazard $\varphi_0(X) > 0$ which depends on a vector $X \in \mathbb{R}^k$ of observable covariates, and a latent random variable $V > 0$ representing unobserved heterogeneity. The latter is assumed to be independent of the covariates X. See Van den Berg (2001) for a review of duration models.

Recall that the hazard function of T conditional on X and V is defined as

$$\theta(t|X, V) = \lim_{\varepsilon \downarrow 0} \frac{\Pr[T \in [t, t+\varepsilon)|T > t, X, V]}{\varepsilon}, \tag{1}$$

which implies that the conditional survival function of T given X and V takes the form

$$S_0(t|X, V) = \Pr[T > t|X, V] = \exp\left(-\int_0^t \theta(\tau|X, V)\, d\tau\right).$$

In the case of the MPH model,

$$\theta(t|X, V) = V.\varphi_0(X).\lambda_0(t), \tag{2}$$

where $\varphi_0(X) > 0$ is the systematic hazard and $\lambda_0(t) \geq 0$ is the baseline hazard, hence

$$S_0(t|X, V) = \exp\left(-V.\varphi_0(X).\int_0^t \lambda_0(\tau)\, d\tau\right) = \exp\left(-V.\varphi_0(X).\Lambda_0(t)\right),$$

where $\Lambda_0(t) = \int_0^t \lambda_0(\tau)\, d\tau$ is the integrated baseline hazard. Of course, in order that $\lim_{t\to\infty} S(t|X, V) = 0$ we need to require that $\lim_{t\to\infty} \Lambda_0(t) = \int_0^\infty \lambda_0(\tau)\, d\tau = \infty$.

Denoting the distribution function of V by $G_0(v)$, the survival function of T conditional on X alone takes the form

$$S_0(t|X) = \int_0^\infty \exp\left(-v.\varphi_0\left(X\right).\Lambda_0\left(t\right)\right) dG_0(v).$$

2.2 Nonparametric identification

Elbers and Ridder (1982) have shown that if $\det\left(\mathrm{Var}\left(X\right)\right) > 0$, (which implies that X does not contain a constant), and

$$E\left[V\right] = \int_0^\infty v dG_0(v) = 1 \tag{3}$$

then the MPH model *without* censoring is nonparametrically identified, in the following sense. If for some distribution function $G(v)$ on $(0, \infty)$ with $\int_0^\infty v dG(v) = 1$, an alternative systematic hazard $\varphi\left(X\right)$ and an alternative integrated baseline hazard $\Lambda(t)$,

$$\int_0^\infty \exp\left(-v\varphi\left(X\right)\Lambda(t)\right) dG(v) = \int_0^\infty \exp\left(-\varphi_0\left(X\right)\Lambda_0(t) dG_0(v)\right)$$

for all $t > 0$ then $\varphi\left(X\right) = \varphi_0\left(X\right)$ a.s., $\Lambda(t) = \Lambda_0(t)$ for all $t > 0$, and $G(v) = G_0(v)$ for all $v > 0$.

Note that the condition (3) is merely a normalization of the general condition that $E\left[V\right] < \infty$.

Heckman and Singer (1984) provide an alternative identification proof based on the results of Kiefer and Wolfowitz (1956), and propose to parametrize G_0 as a discrete distribution.

2.3 Semi-nonparametric specification

Although in principle it is possible to estimate the systematic hazard $\varphi_0\left(X\right)$ nonparametrically, in practice this function is usually parametrized as

$$\varphi_0\left(X\right) = \exp\left(\beta_0' X\right). \tag{4}$$

The same applies to the baseline hazard $\lambda_0(t)$. Also this function can in principle be estimated nonparametrically, but in practice it us usually parametrized. A popular parametrization of $\lambda_0(t)$ is the Weibull specification

$$\begin{aligned} \lambda_0(t) &= \alpha_{01}\alpha_{02}t^{\alpha_{02}-1}, \\ \alpha_0 &= (\alpha_{01}, \alpha_{02})' \in (0, \infty) \times (0, \infty) \end{aligned} \tag{5}$$

with corresponding integrated baseline hazard $\Lambda_0(t) = \int_0^t \lambda_0(\tau)\,d\tau = \alpha_{01}t^{\alpha_{02}}$. However, in this paper I will make no parametric assumptions about the integrated baseline hazard $\Lambda_0(t)$.

Given the parametric specification (4) of the systematic hazard, the conditional survival function takes the form

$$S_0(t|X) = \int_0^\infty \exp\left(-v\exp\left(\beta_0'X\right).\Lambda_0\left(t\right)\right)dG_0\left(v\right),$$

where β_0 is the Euclidean parameter vector and the functions G_0 and $\Lambda_0(t)$ are the non-Euclidean parameters.

An equivalent SNP representation of $S_0(t|X)$ proposed by Bierens (2008a) is

$$S_0(t|X) = H_0\left(\exp\left(-\exp\left(\beta_0'X\right).\Lambda_0\left(t\right)\right)\right) \tag{6}$$

where

$$H_0\left(u\right) = \int_0^\infty u^v dG_0(v). \tag{7}$$

is a distribution function on $[0,1]$, with density

$$h_0(u) = \int_0^\infty vu^{v-1}dG_0(v). \tag{8}$$

Note that the condition (3) is now equivalent to the condition

$$h_0(1) = 1. \tag{9}$$

The nonparametric identification results of Elbers and Ridder (1982) now read as follows. Given an absolutely continuous distribution function $H(u)$ on $[0,1]$ of the form

$$H\left(u\right) = \int_0^\infty u^v dG(v) \tag{10}$$

for some distribution G on $(0,\infty)$, with density $h(u)$ satisfying $h(1) = 1$, and an alternative integrated baseline hazard $\Lambda(t)$, the equality

$$H_0\left(\exp\left(-\exp\left(\beta_0'X\right).\Lambda_0\left(T\right)\right)\right) = H\left(\exp\left(-\exp\left(\beta'X\right).\Lambda\left(T\right)\right)\right) \tag{11}$$

a.s. implies $\beta = \beta_0$, $H(u) = H_0(u)$ and $\Lambda(T) = \Lambda_0(T)$ a.s. It is shown in Bierens (2008a) how to impose the condition $h(1) = 1$. However, the condition that $H(u)$ is of the form (10) is difficult to impose. Without this condition the equality (11) only implies

$$H\left(u\right) = H_0\left(u\right) \text{ for all } u \in \left(\underline{u}, 1\right], \tag{12}$$

where $\underline{u}$ is the lower bound of the support of $U = \exp\left(-\exp(\beta_0'X).\Lambda_0\left(T\right)\right)$.

2.4 Right censoring

Usually the actual duration T is only observed up to an upper bound $\overline{T}$, which may vary per individual. This is called right-censoring. For example, let T be the unemployment spell of a randomly selected worker with an unemployment history. This worker is asked how long he has been unemployed the last time before the interview, and what his characteristics X were at the start of the last unemployment spell. If this worker is no longer unemployed at the time of the interview, his actual unemployment spell T is observed, otherwise let $\overline{T}_0$ be the time between the start of the unemployment spell and the time of the interview. Suppose that in the latter case the worker is interviewed once more a fixed period $\bar{t}$ later. If at the time of the second interview he is no longer unemployed his actual unemployment spell T is observable, otherwise his unemployment spell is right censored with censoring time $\overline{T} = \overline{T}_0 + \bar{t}$. Thus, in this setup, the censoring time $\overline{T}$ has a known lower bound $\bar{t}$. The same applies if the population consists of workers who were unemployed at a particular date, and the sample is drawn $\bar{t}$ periods later.

Although we actually observe $\min\left(T, \overline{T}\right)$, it is advantageous to treat the fixed lower bound $\bar{t}$ as the censoring time because then there is no need to specify the integrated baseline hazard $\Lambda(t)$ parametrically. To see this, let

$$Y = \min\left(T/\bar{t}, 1\right) \tag{13}$$

Then it follows from (6) that

$$\Pr[Y = 1|X] = H_0\left(\exp\left(-\exp\left(\beta_0' X\right).\Lambda_0\left(\bar{t}\right)\right)\right) = H_0\left(\exp\left(-\mu\left(\theta_0, X\right)\right)\right)$$

where $\alpha_0 = \ln\left(\Lambda_0\left(\bar{t}\right)\right)$, $\theta_0 = (\alpha_0, \beta_0')'$ and

$$\mu\left(\theta_0, X\right) = \exp\left(\alpha_0 + \beta_0' X\right),$$

whereas for $\tau \in [0, 1)$,

$$\Pr[Y \leq \tau | X, Y < 1] = \frac{1 - H_0\left(\exp\left(-\mu\left(\theta_0, X\right).F_0\left(\tau\right)\right)\right)}{1 - H_0\left(\exp\left(-\mu\left(\theta_0, X\right)\right)\right)}, \tag{14}$$

where

$$F_0\left(\tau\right) = \Lambda_0\left(\tau.\bar{t}\right)/\Lambda_0\left(\bar{t}\right)$$

is a distribution function on the unit interval $[0, 1]$, with density

$$f_0\left(\tau\right) = \bar{t}.\lambda_0\left(\tau.\bar{t}\right)/\Lambda_0\left(\bar{t}\right). \tag{15}$$

Therefore, the conditional distribution function of Y given X now takes the form

$$\Pr[Y \leq \tau|X] = \Psi\left(\tau|X, \theta_0, F_0, H_0\right) \tag{16}$$

where

$$\Psi\left(\tau|X, \theta, F, H\right) = \begin{cases} 0 & \text{if } \ \tau < 0, \\ 1 - H\left(\exp\left(-\mu\left(\theta, X\right).F\left(\tau\right)\right)\right) & \text{if } \ 0 \leq \tau < 1, \\ 1 & \text{if } \ \tau \geq 1. \end{cases} \tag{17}$$

The distribution functions H_0 and F_0 will be treated as unknown parameters, together with the Euclidean parameter vector θ_0.

The identification conditions and results of Elbers and Ridder (1982) for this case now read as follows.

Theorem 1. *Let the following conditions hold.*
(a) *The distribution function* H_0 *is of the form* $H_0\left(u\right) = \int_0^\infty u^v dG_0\left(v\right)$, *where* $\int_0^\infty v dG_0\left(v\right) = 1$.
(b) *The distribution function* F_0 *is absolutely continuous with density* $f_0(u) > 0$ *for* $u \in (0, 1)$.
(c) *The vector* $X \in \mathbb{R}^k$ *of covariates has a finite and non-singular variance matrix.*
(d) *The distribution function* $\Gamma(w)$ *of* $W = \mu\left(\theta_0, X\right) = \exp\left(\alpha_0 + \beta_0' X\right)$ *satisfies* $\Gamma(w) = \delta_1 \Gamma_c(w) + \delta_2 \Gamma_d(w)$, *where* $\delta_1 \geq 0$, $\delta_2 \geq 0$, $\delta_1 + \delta_2 = 1$, $\Gamma_c(w)$ *is an absolutely continuous distribution function on* $(0, \infty)$ *with support containing an open interval, and* $\Gamma_d(w)$ *is a non-degenerated discrete distribution function.*[1]
(e) *The distribution function of* Y *conditional on* X *takes the form* $\Pr[Y \leq \tau|X] = \Psi\left(\tau|X, \theta_0, H_0, F_0\right)$ *with* $\Psi\left(\tau|X, \theta, H, F\right)$ *defined by* (17).

Suppose that for a $\theta \in \mathbb{R}^{k+1}$, *an absolutely continuous distribution function* F *on* $[0, 1]$ *and an absolutely continuous distribution function* H *on* $[0, 1]$ *with density* h *satisfying* $h(1) = 1$,

$$\Psi\left(Y|X, \theta, F, H\right) = \Psi\left(Y|X, \theta_0, F_0, H_0\right) \ \ a.s. \tag{18}$$

Then $\theta = \theta_0$, $F\left(u\right) = F_0\left(u\right)$ *for all* $u \in [0, 1]$ *and* $H\left(u\right) = H_0\left(u\right)$ *for all* $u \in \left(\underline{u}, 1\right]$, *where* $\underline{u}$ *is the lower bound of the support of* $U = \exp\left(-\exp(W)\right)$.

[1] In general any distribution function can be written as a linear combination of an absolutely continuous distribution function, a discrete distribution function and a singular distribution function. Therefore, this condition is not very restrictive.

Proof: Similar to Bierens (2008b).

2.5 Maximum likelihood

Given a random sample $(Y_1, X_1), ..., (Y_N, X_N)$ from (Y, X) it is possible to construct the log-likelihood function involved. This log-likelihood is now a function of θ and the densities h and f of H and F, respectively, say $\ln(L_N(\theta, f, h))$. It follows from standard maximum likelihood arguments that $E[\ln(L_N(\theta, f, h))] \leq E[\ln(L_N(\theta_0, f_0, h_0))]$ and that under the conditions of Theorem 1,

$$E[\ln(L_N(\theta, f, h))] = E[\ln(L_N(\theta_0, f_0, h_0))]$$

if and only if $\theta = \theta_0$, $f(u) = f_0(u)$ a.e. on $[0, 1]$ and $h(u) = h_0(u)$ a.e. on $(\underline{u}, 1]$. These results suggest to estimate θ_0, f_0 and h_0 by semi-nonparametric maximum likelihood (SNPML). Similar to standard maximum likelihood theory, a crucial condition for the consistency of the SNPML estimators involved is to establish that

$$\sup_{\theta\in\Theta, f\in\mathcal{D}_f, h\in\mathcal{D}_h} \left|N^{-1}\ln(L_N(\theta, f, h)) - N^{-1}E[\ln(L_N(\theta, f, h))]\right| \to 0 \qquad (19)$$

in probability or a.s., where Θ is a compact subset of $\mathbb{R}^{k+1}$ containing θ_0, $\mathcal{D}_f$ is a compact metric space of densities on $[0, 1]$ containing f_0, and $\mathcal{D}_h$ is a compact metric space of densities h on $[0, 1]$ satisfying $h(1) = 1$ and containing h_0.

As mentioned in the conclusion section in Bierens (2008a), the uniform law of large numbers (19) can be established if it is possible to endow the metric spaces $\mathcal{D}_f$ and $\mathcal{D}_h$ with the "sup" metric $\|f_1 - f_2\| = \sup_{0\leq u\leq 1}|f_1(u) - f_2(u)|$. However, it follows from (8) that $h_0(0) = \infty$ if $G_0(1) > 0$, and it follows from (15) that $f_0(0) = \infty$ is possible as well, for example if the baseline hazard $\lambda_0(t)$ is of the Weibull type (5) with $\alpha_{02} \in (0, 1)$. Therefore, it is difficult, if not impossible, to endow the spaces $\mathcal{D}_f$ and $\mathcal{D}_h$ with the "sup" metric.

2.6 Integrated method of moments

In the case of the interval censored MPH model considered in Bierens (2008a) the likelihood function depend on a Euclidean parameter vector and the distribution function H only. Bierens (2008a) constructs a compact metric space

$\mathcal{D}_h$ of density functions h on $[0,1]$ satisfying $h(1)=1$ endowed with the L^1 metric $\int_0^1 |h_1(u)-h_2(u)|du$, so that the space $\mathcal{H}=\left\{H(u)=\int_0^u h(x)\,dx : h\in\mathcal{D}_h\right\}$ of corresponding distribution functions becomes a compact metric space endowed with the sup metric $\sup_{0\leq u\leq 1}|H_1(u)-H_2(u)|$. Along the same lines it is possible to construct a compact metric space $\mathcal{D}_f$ of densities on $[0,1]$ endow with the L^1 metric, so that $\mathcal{F}=\left\{F(u)=\int_0^u f(x)\,dx : f\in\mathcal{D}_f\right\}$ becomes a compact metric space endowed with the sup metric $\sup_{0\leq u\leq 1}|F_1(u)-F_2(u)|$. Therefore, I will propose an alternative to SNPML on the basis of an objective function that depend on θ, F and H only, using the following result.

Lemma 1. *The conditional characteristic function of* $\Psi(Y|X,\theta_0,F_0,H_0)$ *given* X *takes the form* $E\left[\exp\left(i.t.\Psi\left(Y|X,\theta_0,F_0,H_0\right)\right)|X\right]=\varphi\left(t|X,\theta_0,H_0\right)$, *where*

$$\begin{aligned}\varphi\left(t|X,\theta,H\right) &= \exp(i.t)H\left(\exp\left(-\mu\left(\theta,X\right)\right)\right)\\ &\quad+\frac{\sin\left(t.\left(1-H\left(\exp\left(-\mu\left(\theta,X\right)\right)\right)\right)\right)}{t}\\ &\quad+i.\frac{1-\cos\left(t.\left(1-H\left(\exp\left(-\mu\left(\theta,X\right)\right)\right)\right)\right)}{t},\end{aligned}\tag{20}$$

with $i=\sqrt{-1}$. *Consequently,*

$$\Psi\left(Y|X,\theta_0,F_0,H_0\right)\sim I\left(Y=1\right)+I\left(Y<1\right)\left(1-H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right)\right)U,\tag{21}$$

where $U\sim Uniform[0,1]$ *independently of* Y.

Proof: Appendix

As is well-known, characteristic functions of bounded random variables are uniquely determined by their shape in an arbitrary open neighborhood of zero. Therefore, it follows from Theorem 1 and Lemma 1 that

Lemma 2. *If*

$$E\left[\exp\left(i.t.\Psi\left(Y|X,\theta,F,H\right)\right)|X\right]=\varphi\left(t|X,\theta,H\right)\ \text{a.s.}\tag{22}$$

for all t *in an arbitrary open neighborhood of zero then* $\theta=\theta_0$, $F(u)=F_0(u)$ *for all* $u\in[0,1]$ *and* $H(u)=H_0(u)$ *for all* $u\in(\underline{u},1]$.

Proof: Appendix.

Next, let $\Phi : \mathbb{R}^k \to \mathbb{R}^k$ be a bounded one-to-one mapping. It is well-known that conditioning on X is equivalent to conditioning on $\Phi(X)$ because both random vectors generate the same σ-algebra. Therefore, denoting

$$\begin{aligned}\eta(t|Y,X,\theta,F,H) &= t_1.(\varphi(t_1|X,\theta,H) - \exp(i.t_1.\Psi(Y|X,\theta,F,H))) \\ &\quad \times \exp(i.t_2.\Phi(X)) \\ t &= (t_1, t_2') \in \mathbb{R} \times \mathbb{R}^k,\end{aligned}$$

it follows from (22) that $\theta = \theta_0$, $F(u) = F_0(u)$ for all $u \in [0,1]$ and $H(u) = H_0(u)$ for all $u \in (\underline{u}, 1]$ if and only if $E[\eta(t|Y,X,\theta,F,H)] = 0$ for all t in an arbitrary open neighborhood of the origin of $\mathbb{R}^{k+1}$. Consequently,

Lemma 3. *Let*

$$\overline{Q}(\theta,F,H) = \int_T |E[\eta(t|Y,X,\theta,F,H)]|^2 dt, \tag{23}$$

where for example $T = \times_{m=1}^{k+1}[-c_m, c_m]$ *with* $c_m > 0$. *Then* $\overline{Q}(\theta,F,H) = 0$ *if and only if* $\theta = \theta_0$, $F(u) = F_0(u)$ *for all* $u \in [0,1]$ *and* $H(u) = H_0(u)$ *for all* $u \in (\underline{u}, 1]$.

This result suggests to estimate θ_0, F_0 and H_0 by minimizing the objective function

$$\widehat{Q}_N(\theta,F,H) = \int_T \left| \frac{1}{N} \sum_{j=1}^{N} \eta(t|Y_j,X_j,\theta,F,H) \right|^2 dt. \tag{24}$$

How to do that will be discussed below.

3 Sieve estimation

Suppose that H_0 is contained in a compact metric space $\mathcal{H}$ of absolutely continuous distribution functions H on $[0,1]$ with density h satisfying $h(1) = 1$, endowed with the metric $\|H_1 - H_2\| = \sup_{\underline{u} \leq u \leq 1} |H_1(u) - H_2(u)|$, where $\underline{u}$ is defined in Theorem 1. Similarly, let $\mathcal{F}$ be a compact metric space of distribution functions on $[0,1]$ containing F_0, endowed with the metric $\|F_1 - F_2\| = \sup_{0 \leq u \leq 1} |F_1(u) - F_2(u)|$. Moreover, let $\theta_0 \in \Theta$, where Θ is a compact subset of a Euclidean space. How to construct the spaces $\mathcal{H}$ and $\mathcal{F}$ will be discussed in the next section.

For notational convenience, let $\xi_0 = (\theta_0, H_0, F_0)$, $\Xi = \Theta \times \mathcal{H} \times \mathcal{F}$, and combine the Euclidean metric on Θ and the sup metrics on $\mathcal{H}$ and $\mathcal{F}$ in a single metric, for example

$$\begin{aligned} d(\xi_1, \xi_2) &= ||\theta_1 - \theta_2|| + \|H_1 - H_2\| + \|F_1 - F_2\|, \\ \xi_i &= (\theta_i, H_i, F_i),\ i = 1, 2. \end{aligned}$$

Next, let $\overline{Q}(\xi)$ and $\widehat{Q}_N(\xi)$ be defined by (23) and (24), respectively. It is easy to verify from (23) that

$$\overline{Q}(\xi) \text{ is continuous on } \Xi. \tag{25}$$

Moreover, it follows from Lemma 3 that

$$\xi_0 = \arg\min_{\xi \in \Xi} \overline{Q}(\xi) \text{ is unique.} \tag{26}$$

Furthermore, similar to the strong law of large numbers of Jennrich (1969) it can be shown that $\frac{1}{n}\sum_{j=1}^{n} \eta(t|Y_j, X_j, \theta, F, H) \to E[\eta(t|Y, X, \theta, F, H)]$ a.s. uniformly on $T \times \Theta \times \mathcal{H} \times \mathcal{F}$, hence

$$\sup_{\xi \in \Xi} \left| \widehat{Q}_N(\xi) - \overline{Q}(\xi) \right| \to 0 \text{ a.s.} \tag{27}$$

Then, similar to the consistency proof for M-estimators of Euclidean parameter estimators, it follows from (25), (26) and (27) that

$$\widehat{\xi}_N = \arg\min_{\xi \in \Xi} \widehat{Q}_N(\xi) \tag{28}$$

is a strongly consistent estimator of ξ_0, in the sense that $d\left(\widehat{\xi}_N, \xi_0\right) \to 0$ a.s. as $N \to \infty$.

However, in general the computation of $\widehat{\xi}_N$ will not be feasible. The solution to this problem is sieve estimation:

Theorem 2. *Let the conditions* (25), (26) *and* (27) *hold. Suppose it is possible to construct an increasing sequence* Ξ_n, $n = 0, 1, 2, ...$, *of subspaces of* Ξ *such that for each* n *the computation of*

$$\widetilde{\xi}_n = \arg\min_{\xi \in \Xi_n} \widehat{Q}_N(\xi) \tag{29}$$

is feasible. If $\Sigma = \overline{\cup_{n=0}^{\infty}\Sigma_n}$, *where the bar denote the closure, then* $d\left(\widetilde{\xi}_{n_N}, \xi_0\right) \to 0$ a.s. *for any subsequence* n_N *of* N *satisfying* $\lim_{N\to\infty} n_N = \infty$.

Proof: Appendix.

The union $\cup_{n=0}^{\infty}\Sigma_n$ is called the sieve, and the Σ_n's are called the sieve spaces.

4 Densities and Distribution Functions on the Unit Interval

4.1 Series representation

Consider an absolutely continuous distribution function $H(u)$ on the unit interval with density $h(u)$. Since $h(u) = \lim_{n\to\infty} n\left(H(u+n^{-1}) - H(u)\right)$ is a pointwise limit of a sequence of continuous functions, and continuous functions are Borel measurable, it follows that $h(u)$ is a Borel measurable function on $[0,1]$. See for example Bierens (2004, pp. 40-41). Of course, the same applies to general density functions. Moreover, $h(u)$ can be written as $h(u) = f(u)^2$, where $f(u)$ is a Borel measurable function on $[0,1]$ satisfying $\int_0^1 f(u)^2 du = 1$.

To characterize such a function $f(u)$, consider the more general space $L_B^2(0,1)$ of Borel measurable real functions $f(u)$ on $[0,1]$ satisfying $\int_0^1 f(u)^2 du < \infty$. If we endow the space $L_B^2(0,1)$ with the inner product $\langle f_1, f_2\rangle = \int_0^1 f_1(u) f_1(u) du$ and associated norm $||f|| = \sqrt{\langle f, f\rangle}$ and metric $||f_1 - f_2||$ then $L_B^2(0,1)$ becomes a Hilbert space. Recall that a Hilbert space $\mathbb{H}$ is a vector space endowed with an inner product and associated norm and metric such that every Cauchy sequence in $\mathbb{H}$ has a limit in $\mathbb{H}$. See for example Young (1988).

Next, consider an orthonormal sequence $\{\rho_k\}_{k=0}^{\infty}$ in $L_B^2(0,1)$, i.e., $\langle \rho_k, \rho_m\rangle = I(k=m)$. Without loss of generality we may assume that $\rho_0(u) \equiv 1$. The projection of $f \in L_B^2(0,1)$ on $\{\rho_k\}_{k=0}^{n}$ takes the form $f_n(u) = \gamma_0 + \sum_{k=1}^{n} \gamma_k \rho_k(u)$, where $\gamma_k = \langle f, \rho_k\rangle$. The γ_k's are called the Fourier coefficients. They satisfy $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$. Now the orthonormal sequence $\{\rho_k(u)\}_{k=0}^{\infty}$ is complete if for every $f \in L_B^2(0,1)$ with associated projection $f_n(u)$, $\lim_{n\to\infty} ||f - f_n|| =$

0. The latter implies[2] that

$$f(u) = \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \rho_k(u) \text{ a.e. on } [0,1]. \tag{30}$$

Note that $\int_0^1 f(u)^2 du = \sum_{k=0}^{\infty} \gamma_k^2 < \infty$. Therefore, given a complete orthonormal sequence $\{\rho_k\}_{k=0}^{\infty}$ in $L_B^2(0,1)$, every density function $h(u)$ can be written as

$$h(u) = f(u)^2 = \left(\gamma_0 + \sum_{k=1}^{\infty} \gamma_k \rho_k(u)\right)^2, \text{ where } \sum_{k=0}^{\infty} \gamma_k^2 = 1. \tag{31}$$

However, the γ_k's are no longer unique, because we may replace $f(u)$ in (31) by $\phi_B(u) f(u)$ where $\phi_B(u) = I(u \in B) - I(u \in [0,1]\backslash B)$ with B an arbitrary Borel subset of $[0,1]$. Therefore, (31) also holds if we replace γ_k by $\gamma_k^* = \int_0^1 \phi_B(u) f(u) \rho_k(u) du$. In particular, we can choose B such $\gamma_0^* = \int_0^1 \phi_B(u) f(u) du > 0$. Moreover, by Schwarz inequality $\gamma_0^* \leq 1$. Consequently, without loss of generality we may assume that $\gamma_0 \in (0,1]$ in (31). The latter enables us to implement the condition $\sum_{k=0}^{\infty} \gamma_k^2 = 1$ by reparametrizing the γ_k's as

$$\gamma_0 = \frac{1}{\sqrt{1 + \sum_{m=1}^{\infty} \delta_m^2}}, \ \gamma_k = \frac{\delta_k}{\sqrt{1 + \sum_{m=1}^{\infty} \delta_m^2}}, \ k \geq 1,$$

where $\sum_{m=1}^{\infty} \delta_m^2 < \infty$. Thus,

Theorem 3. *Given a complete orthonormal sequence $\{\rho_k\}_{k=0}^{\infty}$ in $L_B^2(0,1)$ with $\rho_0(u) \equiv 1$, for every density function $h(u)$ on $[0,1]$ there exist uncountable many sequences $\{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=1}^{\infty} \delta_m^2 < \infty$ such that*

$$h(u) = \frac{\left(1 + \sum_{m=1}^{\infty} \delta_m \rho_m(u)\right)^2}{1 + \sum_{m=1}^{\infty} \delta_m^2} \text{ a.e.} \tag{32}$$

Moreover, denoting

$$h_0(u) \equiv 1, \ h_n(u) = \frac{\left(1 + \sum_{m=1}^{n} \delta_m \rho_m(u)\right)^2}{1 + \sum_{m=1}^{n} \delta_m^2} \tag{33}$$

[2]In general $\lim_{n\to\infty} ||f - f_n|| = 0$ does not imply $\lim_{n\to\infty} f_n = f$, but in the case of a Hilbert space of functions it does, as $f(u) = \lim_{n\to\infty} \sum_{k=0}^{n} \gamma_k \rho_k(u)$ a.e. See for example Bierens (2009).

for $n \geq 1$, it follows that

$$\int_0^1 |h(u) - h_n(u)|\, du \leq \sqrt{5 \sum_{k=n+1}^{\infty} \delta_k^2} \to 0 \tag{34}$$

as $n \to \infty$.

The result (34) has been proved in Bierens (2009, Theorem 6.1).

Following Gallant and Nychka (1987) the densities $h_n(u)$ will be called SNP densities.

There are many complete orthonormal sequences $\{\rho_k\}_{k=0}^{\infty}$ in $L_B^2(0,1)$. Bierens (2008a) and Bierens and Carvalho (2007) use orthonormal Legendre polynomials. These polynomials can be generated recursively for $n \geq 2$ by

$$\rho_n(u) = \frac{\sqrt{4n^2-1}}{n}(2u-1)\rho_{n-1}(u) - \frac{(n-1)\sqrt{2n+1}}{n\sqrt{2n-3}}\rho_{n-2}(u),$$

starting from $\rho_0(u) = 1$, $\rho_1(u) = \sqrt{3}(2u-1)$. Another complete orthonormal sequence in $L_B^2(0,1)$ is the cosine sequence

$$\rho_0(u) = 1,\ \ \rho_n(u) = \sqrt{2}\cos(n.\pi.u) \text{ for } n \geq 1. \tag{35}$$

4.2 Compact metric spaces of densities and distribution functions

The condition $\sum_{m=1}^{\infty} \delta_m^2 < \infty$ can be imposed by imposing the restrictions $|\delta_k| \leq \overline{\delta}_k$, where $\overline{\delta}_k$ is an a priori chosen positive sequence such that $\sum_{k=1}^{\infty} \overline{\delta}_k^2 < \infty$. For example, let

$$\overline{\delta}_k = \frac{c}{1+\sqrt{k}\ln(k)}, \tag{36}$$

for some constant $c > 0$. It is easy to verify that then $\sum_{k=1}^{\infty} \overline{\delta}_k^2 < c^2 + c^2/\ln(2)$.

These restrictions on the δ_k's also play a key-role in proving compactness:

Theorem 4. *Let $\mathcal{D}$ be the space of densities of the type* (32) *subject to the restrictions $|\delta_k| \leq \overline{\delta}_k$ for some a priori chosen positive sequence $\overline{\delta}_k$ satisfying $\sum_{k=1}^{\infty} \overline{\delta}_k^2 < \infty$, endowed with the L^1 metric $\|h_1 - h_2\|_1 = \int_0^1 |h_1(u) - h_2(u)|\, du$. Then $\mathcal{D}$ is compact. Consequently, the space $\mathcal{C} = \left\{H(u) = \int_0^u h(v)dv,\ h \in \mathcal{D}\right\}$*

endowed with the "sup" metric $\|H_1 - H_2\| = \sup_{0\leq u\leq 1}|H_1(u) - H_2(u)|$ *is compact as well. Moreover, let* $\mathcal{D}_n$ *be the space of SNP densities of the type* (33), *subject to the same restrictions on the* δ_k*'s, and endowed with the same metric as* $\mathcal{D}$*. Then the sequence* $\mathcal{D}_n$ *is dense in* $\mathcal{D}$: $\mathcal{D} = \overline{\cup_{n=0}^{\infty}\mathcal{D}_n}$*. Consequently, the spaces* $\mathcal{C}_n = \left\{H_n(u) = \int_0^u h_n(v)dv,\ h_n \in \mathcal{D}_n\right\}$ *endowed with the sup metric are dense in* $\mathcal{C}$: $\mathcal{C} = \overline{\cup_{n=0}^{\infty}\mathcal{C}_n}$.

Proof: See Bierens (2008a, Theorem 8) for the case where the $\rho_m(u)$ are Legendre polynomials and $\overline{\delta}_k$ is given by (36). However, these results hold for any complete orthonormal sequence $\rho_n(u)$ and any positive sequence $\overline{\delta}_k$ satisfying $\sum_{k=1}^{\infty}\overline{\delta}_k^2 < \infty$.

The parameter space $\mathcal{F}$ for the distribution functions F in Theorem 1 is now equal to the space $\mathcal{C}$, with sieve spaces $\mathcal{F}_n = \mathcal{C}_n$. The same applies to the parameter space $\mathcal{H}$ and sieve spaces $\mathcal{H}_n$ of the distribution functions H in Theorem 1, except that we need to weaken the sup metric of $\mathcal{C}$ to $\|H_1 - H_2\| = \sup_{\overline{u}\leq u\leq 1}|H_1(u) - H_2(u)|$, and impose the conditions

$$1 = h_n(1) = \frac{\left(1 + \sum_{m=1}^{n}\delta_m\rho_m(1)\right)^2}{1 + \sum_{m=1}^{n}\delta_m^2}$$

on $\mathcal{C}_n$ for $n \geq 1$.

If we choose the Legendre polynomials as the complete orthonormal sequence ρ_m then the SNP distribution function $H_n(u) = \int_0^u h_n(v)dv$ has to be computed recursively as well. However, if we choose the cosine series (35) then $H_n(u)$ has a closed form:

$$\begin{aligned}
&H_n(u) = u \\
&+\frac{1}{1 + \sum_{m=1}^{n}\delta_m^2}\left[2\sqrt{2}\sum_{k=1}^{n}\delta_k\frac{\sin(k\pi u)}{k\pi} + \sum_{m=1}^{n}\delta_m^2\frac{\sin(2m\pi u)}{2m\pi}\right. \\
&\left.+2\sum_{k=2}^{n}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin((k+m)\pi u)}{(k+m)\pi} + 2\sum_{k=2}^{n}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin((k-m)\pi u)}{(k-m)\pi}\right]
\end{aligned}$$

See Bierens (2009, Theorem 6.1). Therefore, the cosine series (35) is the most convenient orthonormal sequence for modeling distribution functions on $[0, 1]$.

4.3 What about asymptotic normality?

There is a growing literature on asymptotic normality of the Euclidean part of SNP sieve estimators (the sieve estimator of θ_0 in our case). See Chen (2007) and the references therein. However, these results are based on high-level and difficult to verify assumptions, or the key conditions involved are simply assumed.[3]

Therefore, in practice it is usually assumed that the true non-Euclidean parameters are of the SNP type themselves. In our case the latter amounts to the assumptions that $F_0 \in \cup_{n=0}^{\infty}\mathcal{F}_n$ and $H_0 \in \cup_{n=0}^{\infty}\mathcal{H}_n$, so that there exists a smallest n_1 such that $F_0 \in \mathcal{F}_{n_1}$ and a smallest n_2 such that $H_0 \in \mathcal{H}_{n_2}$. The orders n_1 and n_2 can be estimated consistently using an information criterion similar to the well-known Hannan-Quinn (1979) and Schwarz (1978) information criteria for the dimension of time series models. Given estimators $\widehat{n}_i$ of n_i, $i = 1, 2$, satisfying $\lim_{N\to\infty} \Pr[\widehat{n}_i = n_i] = 1$, one may treat the estimators $\widehat{n}_i$ as the true values. The model then becomes fully parametric, and therefore asymptotic normality can be derived in a standard way, provided that the SNP parameters involved are unique. This is the approach followed by Bierens and Carvalho (2007). As to the uniqueness issue, the problem that the δ_m's in (32) are not unique does not exist for SNP densities based on Legendre polynomials. See Bierens (2008a, Theorem 4). It is my conjecture that this result carries over to SNP densities based on orthonormal cosine functions, but this has not yet been verified.

5 Semi-Nonparametric Estimation of First-Price Auctions

5.1 First-Price Auctions

A first price sealed bids auction (henceforth called *first-price auction*) is an auction where the potential bidder's values for the item to be auctioned off are independent and private, and the bidders are symmetric and risk neutral. The reservation price p_0, if any, is announced in advance and the number $I_0 \geq 2$ of potential bidders is known to each potential bidder.

[3] See for example Theorem 2 and its condition (2.6) in Chen et al. (2003).

The equilibrium bid function of a first-price auction takes the form

$$\beta(v|F_0, I_0) = v - \frac{1}{F_0(v)^{I_0-1}} \int_{p_0}^{v} F_0(x)^{I_0-1} dx \text{ for } v > p_0, \tag{37}$$

if the reservation price p_0 is binding, and

$$\beta(v|F_0, I_0) = v - \frac{1}{F_0(v)^{I_0-1}} \int_{0}^{v} F_0(x)^{I_0-1} dx \tag{38}$$

if the reservation price p_0 is non-binding or absent, where $F_0(v)$ is the value distribution, $I_0 \geq 2$ is the potential number of bidders, and $\underline{v} \geq 0$ is the lower bound of the support of $F_0(v)$. See for example Riley and Samuelson (1981) or Krishna (2002). Thus, if the reservation price p_0 is binding then, with V_j the value for bidder j for the item to be auctioned off, this potential bidder issues a bid $B_j = \beta(V_j|F_0, I_0)$ according to bid function (37) if $V_j > p_0$ and does not issue a bid if $V_j \leq p_0$, whereas if the reservation price p_0 is not binding each potential bidder j issues a bid $B_j = \beta(V_j|F_0, I_0)$ according to bid function (38). In the first-price auction model the individual values V_j, $j = 1, ..., I_0$, are assumed to be independent random drawing from the value distribution F_0. The latter is known to each potential bidder j, and so is the number of potential bidders, I_0.

Bierens and Song (2008a) consider the case where such an auction is independently and identically repeated L times, so that one observes $N = L \times I_0$ independently and identically distributed bids B_j, including possible zero bids if the reservation price is binding. Bierens and Song (2008b) consider the more realistic case where the auctions are heterogeneous, with heterogeneity captured by vectors of auction specific covariates.

5.2 The nonparametric approach

The problem is how to estimate the value distribution F_0 from the observed bids. Guerre et al. (2000) propose an indirect nonparametric kernel estimation approach, for the case similar to the one in Bierens and Song (2008a). In the case that the reservation price is nonbinding their approach is based on the inverse bid function $v = b + (I_0 - 1)^{-1} \Lambda(b)/\lambda(b)$, where I_0 is the number of bidders, v is a private value, b is a corresponding bid, and $\Lambda(b)$ is the distribution function of bids with density $\lambda(b)$. The latter two functions are estimated via nonparametric kernel methods, as $\hat{\Lambda}(b)$ and $\hat{\lambda}(b)$, respectively.

Using the pseudo-private values $\widetilde{V} = B + (I_0 - 1)^{-1} \hat{\Lambda}(B)/\hat{\lambda}(B)$, where each B is an observed bid, the density of the private value distribution can now be estimated by kernel density estimation. However, the ratio $\hat{\Lambda}(b)/\hat{\lambda}(b)$ may be an unreliable estimate of $\Lambda(b)/\lambda(b)$ near the boundary of the support of $\lambda(b)$. To solve this problem, Guerre et al. (2000) use a trimming procedure which amounts to discarding pseudo-private values $\widetilde{V}$ corresponding to bids B that are too close to the boundary of the (known) support of the bid distribution.

5.3 Integrated simulated moments sieve estimation

Bierens and Song (2008a) take a different route, as follows. Suppose that there is no reservation price, so that the number of potential bidders I_0 is observable. The data consist of N bids B_j generated according to

$$B_j = V_j - \frac{1}{F_0(V_j)^{I_0 - 1}} \int_0^{V_j} F_0(x)^{I_0 - 1} dx,$$

where the V_j's are random drawings from the distribution F_0. It can be shown that if $E[V_j] < \infty$ then the bid distribution has bounded support.

Given a candidate value distribution F, Bierens and Song (2008a) propose to draw a random sample $\widetilde{V}_1(F), ..., \widetilde{V}_N(F)$ from F,[4] generate simulated bids $\widetilde{B}_j(F)$ according to

$$\widetilde{B}_j(F) = \widetilde{V}_j(F) - \frac{1}{F\left(\widetilde{V}_j(F)\right)^{I_0 - 1}} \int_0^{\widetilde{V}_j(F)} F(x)^{I_0 - 1} dx,$$

and then form the objective function

$$\widehat{Q}_N(F) = \int_T \left| \frac{1}{N} \sum_{j=1}^{N} \left(\exp(i.t.B_j) - \exp\left(i.t.\widetilde{B}_j(F)\right) \right) \right|^2 dt,$$

where $T = \left[-\bar{t}, \bar{t}\right]$ for some $\bar{t} > 0$. Note that pointwise in F, $\widehat{Q}_N(F) \to \overline{Q}(F)$ a.s., where

$$\overline{Q}(F) = \int_T \left| E\left[\exp(i.t.B_j) - \exp\left(i.t.\widetilde{B}_j(F)\right)\right] \right|^2 dt.$$

[4] By drawing a random sample $U_1, ..., U_N$ from the uniform $[0,1]$ distribution and solving $F\left(\widetilde{V}_j(F)\right) = U_j$ for each F and j.

Since the B_j's are bounded, it follows that $\overline{Q}(F) = 0$ if and only if $F = F_0$, similar to the MPH model case.

To construct a compact metric space $\mathcal{F}$ of absolutely continuous distribution functions on $(0, \infty)$, Bierens (2008a) observes that given an a priori chosen absolutely continuous distribution function G with support $(0, \infty)$ and density g, any absolutely continuous distribution function F on $(0, \infty)$ can be written as $F(v) = H(G(v))$ with density $f(v) = h(G(v))\,g(v)$, where $H(u) = F(G^{-1}(u))$ is a distribution function on $[0, 1]$ with density $h(u)$. Therefore, for a given distribution function G and with $\mathcal{C}$ the compact metric space defined in Theorem 4,

$$\mathcal{F} = \{F(v) = H(G(v)) : H \in \mathcal{C}\}$$

is a compact metric space of distribution functions on $(0, \infty)$ endowed with the sup metric $||F_1 - F_2|| = \sup_{v>0} |F_1(v) - F_2(v)|$, with corresponding sieve spaces

$$\mathcal{F}_n = \{F(v) = H(G(v)) : H \in \mathcal{C}_n\}.$$

All the conditions of Theorem 2 hold for this case. Therefore, the sieve estimator

$$\widetilde{F}_{n_N} = \arg\min_{F \in \mathcal{F}_{n_N}} \widehat{Q}_N(F)$$

is strongly consistent, uniformly on $(0, \infty)$, i.e.,

$$\sup_{v>0} |\widetilde{F}_{n_N}(v) - F_0(v)| \to 0 \text{ a.s.}$$

6 Appendix

6.1 Proof of Lemma 1

We can write

$$\begin{aligned} \Psi(Y|X, \theta_0, F_0, H_0) &= I(Y = 1) \\ &+ I(Y < 1)\left(1 - H_0\left(\exp\left(-\mu\left(\theta_0, X\right)\right)\right)\right).U \end{aligned} \tag{39}$$

where

$$U = \frac{1 - H_0\left(\exp\left(-\mu\left(\theta_0, X\right).F_0\left(Y\right)\right)\right)}{1 - H_0\left(\exp\left(-\mu\left(\theta_0, X\right)\right)\right)}, \tag{40}$$

hence

$$\begin{aligned}
&\varphi\left(t|X,\theta_0,F_0,H_0\right)=\exp(i.t)E\left[I(Y=1)|X\right] \\
&+E\left[I(Y<1)\exp\left(i.t\left(1-H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right)\right)U\right)|X\right] \\
&=\exp(i.t)H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right) \\
&+E\left[I(Y<1)E\left(\exp\left(i.t\left(1-H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right)U\right)\right)|X,Y<1\right)|X\right]
\end{aligned} \tag{41}$$

As is well known, if we insert a random drawing from a univariate continuous (conditional) distribution in its own (conditional) distribution function the result is a random drawing from the uniform $[0,1]$ distribution. The conditional distribution (14) is such a distribution, hence $\Pr\left[U\leq u|X,Y<1\right]=u$ for $u\in[0,1]$, and therefore

$$E\left[\exp\left(i.t.U\right)|X,Y<1\right]=\frac{\sin(t)}{t}+i\frac{1-\cos(t)}{t}.$$

Replacing t by $t\left(1-H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right)\right)$ it follows now that

$$\begin{aligned}
&E\left(\exp\left(i.t\left(1-H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right)\right).U\right)|X,Y<1\right) \\
=\;&\frac{\sin\left(t\left(1-H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right)\right)\right)}{t\left(1-H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right)\right)} \\
&+i\frac{1-\cos\left(t\left(1-H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right)\right)\right)}{t\left(1-H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right)\right)}.
\end{aligned} \tag{42}$$

The result (20) of Lemma 1 follows now straightforwardly from (41) and (42). The result (21) follows from the fact that the two sides of (21) have the same characteristic function.

6.2 Proof of Lemma 2

Similar to (39) and (40) we can write

$$\Psi\left(Y|X,\theta,F,H\right)=I(Y=1)+I(Y<1)\left(1-H\left(\exp\left(-\mu\left(\theta,X\right)\right)\right)\right)\widetilde{U}$$

where

$$\widetilde{U}=\frac{1-H\left(\exp\left(-\mu\left(\theta,X\right).F\left(Y\right)\right)\right)}{1-H\left(\exp\left(-\mu\left(\theta,X\right)\right)\right)}.$$

Hence,

$$\begin{aligned}
&E\left[\exp\left(i.t.\Psi\left(Y|X,\theta,F,H\right)\right)|X\right]=\exp(i.t)H_0\left(\exp\left(-\mu\left(\theta_0,X\right)\right)\right) \\
&+E\left[I(Y<1)E\left(\exp\left(i.t\left(1-H\left(\exp\left(-\mu\left(\theta,X\right)\right)\right)\right)\widetilde{U}\right)|X,Y<1\right)|X\right]
\end{aligned}$$

Moreover, $\varphi(t|X,\theta,H)$ can be written as

$$\begin{aligned}\varphi(t|X,\theta,H) &= \exp(i.t)H(\exp(-\mu(\theta,X)))+(1-H(\exp(-\mu(\theta,X)))) \\ &\times E[\exp(i.t.(1-H(\exp(-\mu(\theta,X))))U)|X]\end{aligned}$$

where U is a random drawing from the uniform $[0,1]$ distribution.

Given that the equality (22) holds in an open neighborhood of zero it follows that for $m=1,2,3,....$

$$\begin{aligned}&H_0(\exp(-\mu(\theta_0,X))) \\ &+E\left[I(Y<1)E\left(\left(1-H(\exp(-\mu(\theta,X))).\widetilde{U}\right)^m|X,Y<1\right)|X\right] \\ &=\left.\frac{\partial^m E[\exp(i.t.\Psi(Y|X,\theta,F,H))|X]}{i^m(\partial t)^2}\right|_{t=0}=\left.\frac{\partial^m\varphi(t|X,\theta,H)}{i^m(\partial t)^2}\right|_{t=0} \\ &=H(\exp(-\mu(\theta,X))) \\ &+(1-H(\exp(-\mu(\theta,X))))E[((1-H(\exp(-\mu(\theta,X))))U)^m|X]\end{aligned}$$

Letting $m\to\infty$ yields

$$H(\exp(-\mu(\theta,X)))=H_0(\exp(-\mu(\theta_0,X))).$$

It follows now straightforwardly that $\varphi(t|X,\theta,H)=\varphi(t|X,\theta_0,H_0)$ a.s. for all $t\in\mathbb{R}$, hence

$$\Psi(Y|X,\theta,F,H)\sim\Psi(Y|X,\theta_0,F_0,H_0). \tag{43}$$

Finally, it is not too hard to verify that (43) is only possible if the equality (18) in Theorem 1 holds, so that the result of Lemma 2 follows from Theorem 1.

6.3 Proof of Theorem 2

For $n\geq 1$, let $\xi_n=\arg\min_{\xi\in\Xi_n}\overline{Q}(\xi)$. If $\xi_0\in\cup_{n=0}^{\infty}\Sigma_n$ then $\xi_0\in\Xi_{\overline{n}}$ for some fixed $\overline{n}$, hence $\overline{Q}(\xi_{n_N})=\overline{Q}(\xi_0)$ once $n_N\geq\overline{n}$. Next, suppose that $\xi_0\in\overline{\cup_{n=0}^{\infty}\Sigma_n}\backslash\cup_{n=0}^{\infty}\Sigma_n$. Then for every $\varepsilon>0$ there exists an element $\xi_*\in\cup_{n=0}^{\infty}\Sigma_n$ such that $d(\xi_0,\xi_*)<\varepsilon$, hence

$$\limsup_{N\to\infty}\left(\overline{Q}(\xi_{n_N})-\overline{Q}(\xi_0)\right)\leq\overline{Q}(\xi_*)-\overline{Q}(\xi_0)\leq\sup_{d(\xi_0,\xi)<\varepsilon}\left|\overline{Q}(\xi)-\overline{Q}(\xi_0)\right|$$

Letting $\varepsilon \downarrow 0$ it follows by continuity that $\lim_{N\to\infty} \overline{Q}(\xi_{n_N}) = \overline{Q}(\xi_0)$. It follows now from the easy inequality

$$0 \leq \overline{Q}\left(\widehat{\xi}_{n_N}\right) - \overline{Q}(\xi_0) \leq 2.\sup_{\xi\in\Xi}\left|\widehat{Q}_N(\xi) - \overline{Q}(\xi)\right| + \overline{Q}(\xi_{n_N}) - \overline{Q}(\xi_0)$$

that $\overline{Q}\left(\widehat{\xi}_{n_N}\right) \to \overline{Q}(\xi_0)$ a.s. Due to the continuity of $\overline{Q}(\xi)$ and the uniqueness of ξ_0 the latter implies that $d\left(\widehat{\xi}_{n_N}, \xi_0\right) \to 0$ a.s.[5]

References

Bierens, H. J., 2004, *Introduction to the Mathematical and Statistical Foundations of Econometrics*, Cambridge University Press.

Bierens, H. J., 2008a, Semi-Nonparametric Interval Censored Mixed Proportional Hazard Models: Identification and Consistency Results, *Econometric Theory,* 24, 749-794.

Bierens, H. J., 2008b, Semi-Nonparametric Identification of the Right Censored Mixed Proportional Hazard Model, Lecture note (http://econ.la.psu.edu/~hbierens/IMPHM.PDF).

Bierens, H.J., 2009, Hilbert Space Theory and Its Applications in Econometrics, Lecture note (http://econ.la.psu.edu/~hbierens/HILBERTSPACE .PDF).

Bierens, H.J. & J. R. Carvalho, 2007, Semi-Nonparametric Competing Risks Analysis of Recidivism, *Journal of Applied Econometrics,* 22, 971-993.

Bierens, H. J. & H. Song, 2008a, Semi-Nonparametric Estimation of Independently and Identically Repeated First-Price Auctions via an Integrated Simulated Moments Method, Working paper (http://econ.la.psu.edu/~hbierens /AUCTIONS_IID.PDF).

Bierens, H. J. & H. Song, 2008b, Semi-Nonparametric Estimation of First-Price Auctions Models with Auction-Specific Heterogeneity via an Integrated Simulated Conditional Moments Method, Working paper (http://econ.la.psu .edu/~hbierens/AUCTIONS_HETERO.PDF).

Chen, X., 2007, Large Sample Sieve Estimation of Semi-Nonparametric Models. In J. Heckman & E. Leamer (eds.), *Handbook of Econometrics, Vol. 6,* Ch. 76, Elsevier (in press).

[5] Similar to Jennrich's (1969) consistency proof of nonlinear least squares estimators.

Chen, X., O. Linton & I. Van Keilegom, 2003, Estimation of Semiparametric Models when the Criterion Function is Not Smooth, *Econometrica*, 71, 1591-1608.

Eastwood, B. J. & A. R. Gallant, 1991, Adaptive Rules for Semi-Nonparametric Estimators that Achieve Asymptotic Normality, *Econometric Theory,* 7, 307-340.

Elbers, C. & G. Ridder, 1982, True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model, *Review of Economic Studies,* 49, 403-409.

Gallant, A. R., 1981, On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form, *Journal of Econometrics,* 15, 211-245.

Gallant, A. R. & D. W. Nychka, 1987, Semi-Nonparametric Maximum Likelihood Estimation, *Econometrica,* 55, 363-390.

Guerre, E., I, Perrigne & Q. Vuong, 2000, Optimal Nonparametric Estimation of First-Price Auctions, *Econometrica,* 68, 525-574.

Hannan, E.J., & B.G. Quinn, 1979, The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society, Series B,* 41, 190-195.

Heckman, J. J., 1979, Sample Selection Bias as a Specification Error, *Econometrica,* 47, 153-161.

Heckman, J. J. & B. Singer, 1984, A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data, *Econometrica*, 52, 271-320.

Jennrich, R. I., 1969, Asymptotic Properties of Nonlinear Least Squares Estimators, *Annals of Mathematical Statistics,* 40, 633-643.

Kiefer, J. & J. Wolfowitz, 1956, Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters, *The Annals of Mathematical Statistic*, 27, 887-906.

Krishna, V., 2002, *Auction Theory*, Academic Press.

Kronmal, R. & M. Tarter, 1968, The Estimation of Densities and Cumulatives by Fourier Series Methods, *Journal of the American Statistical Association,* 63, 925-952.

Lancaster, T., 1979, Econometric Methods for the Duration of Unemployment, *Econometrica*, 47, 939-956.

Riley, G.J. & W.F. Samuelson, 1981, Optimal Auctions, *American Economic Review,* 71, 381-392.

Schwarz, G., 1978, Estimating the Dimension of a Model, *Annals of Statistics,* 6, 461-464.

Van den Berg, G. J., 2001, Duration models: Specification, identification, and multiple durations. In J. Heckman & E. Leamer (eds.), *Handbook of Econometrics, Vol. 5,* Ch. 55, pp. 3381-3460, Elsevier.

Young, N., 1988, *An Introduction to Hilbert Space*, Cambridge University Press.